

BAB 3 METODE PENELITIAN

3.1 Pengumpulan Data (*Data gathering*).

Pengumpulan data harus mampu mendeskripsikan data yang ada, serta memiliki kontribusi terhadap pengetahuan. Data yang tidak lengkap perlu ditangani secara sistematis agar tidak menimbulkan kesalahan pada saat pengolahan data. Jika memang tidak dapat disesuaikan maka data tersebut bisa dihilangkan.

Data yang tidak konsisten muncul karena adanya perubahan aplikasi pengisi data atau migrasi sistem. Sedangkan data yang tidak rapi biasanya merupakan data dengan nilai aneh yang perlu dikonversi agar dapat digunakan kembali.

Pengumpulan data pada penelitian ini diambil dari web *Yahoo Finance* yaitu;

1. *Tata Consultancy Service* (TCS) sebanyak 456 data saham dan;
2. *Indonesian Stock Exchange* (Bursa Efek Indonesia/ IDX) 440 data saham.

Periode pengambilan data diambil mulai dari untuk *Tata Consultancy Service* (TCS) dan IDX mulai tanggal 1 Januari 2014 hingga 30 November 2015 . Pada proses pengolahan dataset menggunakan ANN, data saham diolah menggunakan 2 (dua) metode, yakni masing-masing dataset dibagi menjadi 2 grup, yaitu 70% untuk data *training* dan 30% untuk data *testing* [40], Dan metode kedua menggunakan pengolahan dataset secara keseluruhan tanpa membagi dataset saham ke dalam data *training* dan data *testing*.

Asumsi yang mendasari bahwa pengolahan dataset secara keseluruhan di dasarkan pada *10-fold cross-validation* adalah pilihan terbaik untuk mendapatkan hasil validasi yang akurat, karena *10-fold cross-validation* akan mengulang pengujian sebanyak 10 kali dan hasil pengukuran adalah nilai rata-

rata dari 10 kali pengujian [39]. Guna membedakan proses eksperimen pada penelitian ini membedakan proses yang membagi dataset kedalam 70% data training dan 30% dalam data testing disebut dengan proses *split test validation*. Nilai RMSE yang dihasilkan dari proses *split test validation* disebut dengan *RMSE Testing*, dan proses yang menggunakan data keseluruhan tanpa dibagi 70% dataset sebagai data training dan 30% data testing disebut *10 fold cross validation* dan RMSE yang dihasilkan disebut dengan *RMSE Validasi*.

Pengolahan menggunakan ANN untuk memperoleh nilai RMSE terdiri data saham yang memiliki 7 (tujuh) atribut, antara lain :

1. data tanggal (*price date*),
2. data nilai pembukaan (*open price*),
3. data nilai tertinggi (*high price*),
4. data nilai terendah (*low price*),
5. data perubahan nilai (*traded value*) / *adjusted value*,
6. data jumlah nilai yang berubah (*number off trades*) / (*traded quantity*) dan
7. data nilai penutupan (*close price*) t.
8. label dataset adalah nilai penutupan $t-1$.

Penggunaan data tanggal (*price date*), sebagai atribut dalam penelitian saham berguna untuk menentukan waktu yang paling tepat untuk melakukan *short selling* (penjualan jangka pendek yang menampilkan nilai riil dari saham tersebut), dan *long selling* (kenampakan nilai *trend* / pola nilai saham secara riil pada kurva perdagangan saham) [7].

Prediksi saham dengan menggunakan metode konvensional seperti *technical analisis*, *fundamental analisis* atau gabungan keduanya lebih menekankan pada perubahan *trend* pola *pattern* yang dihasilkan oleh data nilai pembukaan (*open price*) yaitu harga saham pada awal pembukaan perdagangan dan data nilai

penutupan (*close price*) t merupakan harga pada akhir periode. Sementara data nilai tertinggi (*high price*), data nilai terendah (*low price*) mengacu pada harga tertinggi dan terendah yang diperoleh untuk saham selama periode tersebut.

Keempat atribut ini juga merupakan atribut yang paling umum digunakan dalam penelitian prediksi saham dengan menggunakan metode *machine learning*, terutama metode *Artificial Neural Network* (ANN) [29]. Akurasi penelitian dengan metode ANN ditentukan oleh bentuk arsitektur jaringan yang dihasilkan dari penelitian nilai historis dari keempat atribut yang digunakan tersebut.

Investor di pasar saham senantiasa ingin memaksimalkan keuntungan dengan membeli atau menjual investasi mereka pada waktu yang tepat. Investasi ini pun meliputi nilai saham. Karena data pasar memiliki pola non linier yang dapat berubah sewaktu-waktu [41]. Menganalisis data pasar saham untuk memprediksi nilai saham dimasa yang akan datang menjadi objek penelitian yang lebih menantang. Pengembangan penggunaan ANN diperlukan untuk memperkuat akurasi prediksi yang dihasilkan. Oleh karena itu, pada penelitian ini menggunakan atribut data perubahan nilai (*traded value*)/ *adjusted value*, dan data jumlah nilai yang berubah (*number off trades*)/ (*traded quantity*) sebagai atribut yang membedakan penelitian ini dengan penelitian ANN yang telah ada sebelumnya.

Penentuan kerangka/ lingkup waktu prediksi merujuk pada Larsen [42], terbagi menjadi 4 (empat) kriteria waktu antara lain:

1. Posisi perdagangan (*Position trades*); perdagangan saham dapat dilakukan dari minggu ke bulan.
2. Ayunan perdagangan (*Swing trades*); perdagangan saham dilakukan dalam waktu dua sampai lima hari.

3. Hari perdagangan (*Day trades*); saham yang dibeli dan dijual dalam hari yang sama.
4. Momentum perdagangan (*Momentum trades*); saham yang dijual dan dibeli dalam hitungan detik, menit atau jam.

Setiap kerangka waktu prediksi memiliki kelebihan dan kekurangan sesuai dengan tujuan dilakukannya prediksi. Pada penelitian ini label dataset dari y adalah nilai penutupan $t-1$. Nilai $t-1$ ini menunjukkan prediksi didasarkan pada 1 (satu) hari setelah nilai penutupan. Penentuan label ini berdasarkan analisis fundamental dari harga saham yang menyusun data historis nilai saham dataset TCS (*Tata Consultancy Service*) dan IDX (*Indonesian Exchange Rate*) yang menetapkan nilai historis saham berdasarkan 5 (lima) hari/ minggu (*weekly*) dikombinasikan dengan penggunaan metode *windowing* pada proses awal pengolahan data (*pre-processing*).

Pengolahan data awal yang dilakukan dengan menggunakan teknik *windowing* diterapkan untuk seluruh dataset data *univariate* dan *multivariate*. Data *univariate* adalah data tunggal yang digunakan dalam proses prediksi. Pada penelitian ini, data tunggal yang digunakan adalah nilai penutupan (*close price*). Penggunaan nilai penutupan didasarkan pada dua metode penentuan nilai prediksi secara umum dalam teori ekonomi. Jika nilai prediksi lebih kecil dibandingkan nilai penutupan aktual maka, kondisi tersebut dalam ekonomi nilai saham sedang berada pada *trend* naik, sebaliknya jika nilai prediksi lebih kecil dari nilai penutupan aktual maka kondisi tersebut dinamakan saham sedang berada pada *trend* turun.

Jumlah *window step* untuk data *univariate* yang digunakan adalah 5 (lima), sesuai dengan pola historis dari sementara untuk data *multivariate* menggunakan *window step* sebanyak 1 (satu). Perbedaan jumlah *window step*

bertujuan untuk membentuk arsitektur jaringan sama antara *univariate* dan *multivariate* sehingga mudah dikomparasi. Target prediksi adalah untuk memprediksi nilai saham untuk bulan oktober hingga desember 2015. Tiga tabel berikut menampilkan sebagian dari dataset saham yang digunakan dalam penelitian. Tabel 3.1 untuk dataset TCS (*Tata Consultancy Service*) dan Tabel 3.2 untuk dataset IDX (*Indonesia Stock Exchange*).

Tabel 3-1 Sebagian data saham TCS

<i>PriceDate</i>	<i>OpenPrice</i>	<i>Highprice</i>	<i>lowprice</i>	<i>closeprice</i>	<i>tradedvalue</i>	<i>Numberoftrades</i>	<i>TradedQuantity</i>
4/18/2012	1100	1113.8	1093.05	1096.8	105168282	5179	95376
4/19/2012	1096.65	1104.95	1091.35	1102.2	75084397	3204	68253
4/20/2012	1099.7	1107.5	1078	1089.4	101559493	4346	92808
4/23/2012	1082	1102.45	1046.55	1059.25	190994729	7448	177800
4/24/2012	1120	1229.75	1120	1195.25	2107821943	60741	1794158

Tabel 3-2 Sebagian data saham IDX

<i>Date</i>	<i>Open</i>	<i>High</i>	<i>Low</i>	<i>Close</i>	<i>Volume</i>	<i>Adj Close</i>
11/7/2014	24.6	24.74	24.49	24.73	34900	24.73
11/6/2014	24.81	24.83	24.61	24.62	77900	24.62
11/5/2014	24.99	25.01	24.64	24.86	30000	24.86
11/4/2014	25.08	25.09	24.81	25.01	55500	25.01

3.2 Pengolahan Awal Data (Data pre-processing).

Sebagai tindak lanjut dari pengumpulan data, tahap ini terdiri dari pengolahan data awal saham untuk menentukan atribut ANN untuk *univariate* dan *multivariate*.

a. Penentuan *input* dan *output*

Data *input* dan *output* dari penelitian ini adalah data *time series*. Data *time series* ini dibagi ke dalam 5 *window size*, yang mengakibatkan penambahan jumlah atribut baru pada dataset, sehingga atribut yang

digunakan adalah seperti pada tabel 3-3. *Output (label)* pada eksperimen ini adalah nilai penutupan untuk 1 (satu) hari berikutnya.

Tabel 3-3 Tabel Atribut Baru

Tanggal-4 _(t-5)	Nilai Penutupan-4 _(t-5)
Tanggal-3 _(t-4)	Nilai Penutupan-4 _(t-4)
Tanggal-2 _(t-3)	Nilai Penutupan-4 _(t-3)
Tanggal-1 _(t-2)	Nilai Penutupan-4 _(t-2)
Tanggal-0 _(t-1)	Nilai Penutupan 4 _(t-1)

Data *input* dan *output* dari penelitian ini adalah data *time series*, dimana yang menjadi *input* lain data tanggal (*Price Date*), data nilai pembukaan (*Open Price*), data nilai tertinggi (*High Price*), data nilai terendah (*Low Price*), data nilai penutupan (*Close Price*), data perubahan nilai (*Adjusted Close*), data jumlah nilai yang berubah (*Number Of Trades*).

Input data ini digunakan sebagai data *univariate* hanya menggunakan data tanggal (*Price Date*) sebagai id dan data nilai penutupan (*Close Price*) sebagai atribut. Sedangkan label data *univariate* akan muncul dari proses *windowing* yang dilakukan yaitu nilai penutupan $t-1$ (*Close Price* $t-1$).

Seluruh atribut input data *multivariate* digunakan dalam penelitian. Proses awal data *multivariate* juga menggunakan teknik *windowing*, sehingga label dari pengolahan data sama dengan data *univariate* yaitu nilai penutupan $t-1$ (*Close Price* $t-1$). Proses *windowing* ini dilakukan dengan tujuan membantu peneliti mengamati pola *trend* perubahan data yang terjadi, seperti dilakukan oleh Z. Xie, M. Ward, and E. Rundensteiner [43].

Proses perdagangan saham dibursa efek saham secara *on line* maupun *off line*, dilakukan dalam waktu 5 (lima) hari dalam satu minggu, sehingga

menimbulkan adanya data saham kosong pada tanggal yang tidak terekam perlakuan untuk data tersebut adalah dengan menggunakan nilai rata-rata (*mean*) dari harga penutupan setiap minggunya. Dengan tidak adanya data yang terlewat maka, dataset saham telah dapat digunakan dalam penelitian dengan pendekatan *time series*.

b. Normalisasi

Sebelum data input dan label diimplementasikan ke dalam *neural network*, terlebih dahulu harus melalui *preprocessing data* berupa penskalaan atau normalisasi data. Tujuannya agar *neural network* dapat mengenali data yang akan menjadi masukan bobot-bobotnya. Normalisasi data ini berada dalam *range* 0 hingga 1. Rumus normalisasi adalah sebagai berikut :

$$= \frac{D - D_{min}}{D_{max} - D_{min}} \quad (3.1)$$

dimana N = data yang telah dinormalisasi

D = data yang akan dinormalisasi

D_{min} = data terkecil dalam sebuah *range*

D_{max} = data terbesar dalam sebuah *range*

3.3 Eksperimen, Pengujian Metode dan Evaluasi Hasil

Tahapan-tahapan eksperimen ANN pada penelitian ini adalah sebagai berikut :

1. Tentukan struktur/arsitektur dari *backpropagation neural network* (BPNN).
2. Untuk mendapatkan nilai RMSE yang kecil, diperlukan arsitektur *neural network* yang tepat. Oleh karena itu diperlukan pengaturan pada attribute parameter-parameter *neural network*, seperti *training cycles*, *learning rate*, *momentum* dan *hidden layer*. Penentuan *neural network*, seperti *training cycles*, *learning rate*, *momentum* dan *hidden layer* dalam penelitian ini menggunakan *trial and error*. *Training cycles* yang optimal selanjutnya

digunakan untuk menentukan *learning rate* dan *momentum* yang optimal. Nilai *training cycles* dan *learning rate* yang optimal selanjutnya digunakan untuk menentukan *momentum* yang optimal yang paling baik.

3. Memasukkan dataset nilai saham yang diambil dari database *Yahoo Finance* yaitu *Tata Consultancy Service (TCS)* dan *IDX (Indonesian Stock Exchange)* yang memiliki 7 atribut dan 1 label.
4. Pengolahan data (*data preprocessing*).
5. Pada pengolahan data awal dilakukan normalisasi dengan merubah data sehingga data berada dalam skala antara -1 dan 1. Selanjutnya dilakukan proses yang disebut dengan data *cleaning*, yang berfungsi untuk membersihkan nilai yang kosong dan tidak konsisten (*missing value dan noisy*).
6. Lakukan *training* pada *neural network* untuk menghitung *network error* dari *output* yang dihasilkan.
7. Cek apakah kondisi sudah terpenuhi. Kondisi dinyatakan terpenuhi jika solusi sekarang sudah konvergen (posisi semua *neuron* menuju ke satu nilai yang sama atau telah mencapai iterasi yang ditentukan).
8. Jika belum terpenuhi, maka kembali ke langkah (3) dengan iterasi yang diperbarui menjadi $i = i+1$.
9. Jika kondisi telah terpenuhi, maka tahapan berikutnya adalah tahap evaluasi. Pada tahap ini dibahas tentang hasil evaluasi dari eksperimen yang telah dilakukan. Pengukuran kinerja dilakukan dengan menghitung rata-rata *error* yang terjadi melalui nilai RMSE yang diperoleh. Semakin kecil nilai RMSE yang diperoleh menunjukkan semakin dekat nilai prediksi dengan nilai sebenarnya. RMSE dicari dengan menggunakan rumus pada rumus 2.15.
10. Jika nilai RMSE telah diperoleh, maka BPNN yang digunakan untuk prediksi nilai saham telah terpenuhi.

11. Untuk mengetahui nilai asli, maka perlu dilakukan langkah denormalisasi seperti persamaan di bawah ini.

$$= * (-) + \quad (3.3)$$

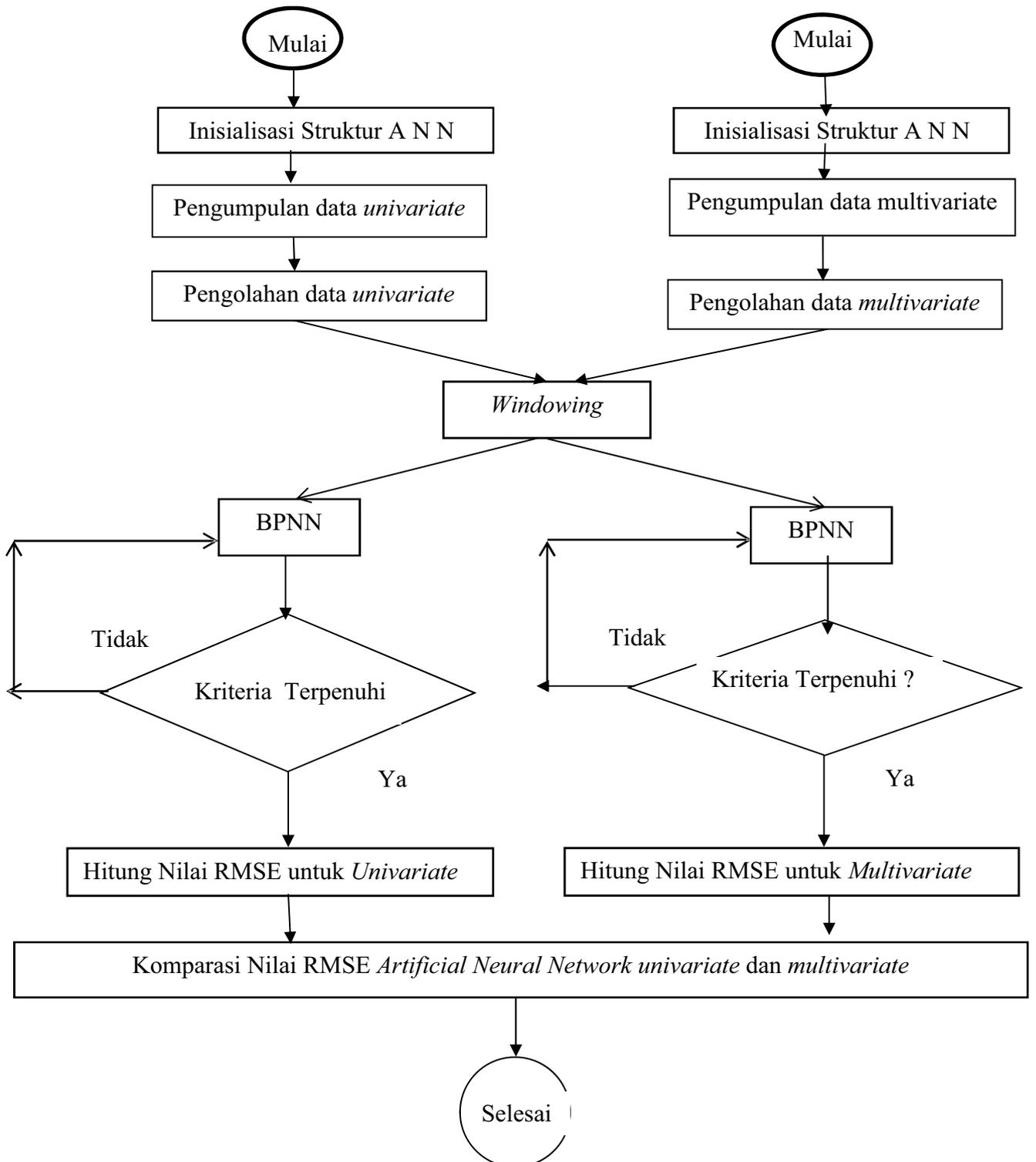
dimana D : data yang sudah didenormalisasi

O : nilai output yang dihasilkan

D_{maks} : data terbesar dalam sebuah *range*

D_{min} : data terkecil dalam sebuah *range*

Tahapan dari eksperimen tersebut selengkapnya dapat dilihat pada Gambar 3-1



Gambar 3-1 Tahapan Eksperimen Komparasi Prediksi Nilai Saham